

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2002, 14:26:09 ; Search time 2530 Seconds  
(without alignments)  
18116.861 Million cell updates/sec

Title: US-09-635-501-1  
Perfect score: 3396  
Sequence: 1 gaattcggttcacataa.....aaaaaaaaagggcgccgc 3396

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues  
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

1:	em_estba:*
2:	em_esthum:*
3:	em_estin:*
4:	em_estmu:*
5:	em_estov:*
6:	em_estpl:*
7:	em_estro:*
8:	em_htc:*
9:	gb_estl:*
10:	gb_est2:*
11:	gb_htc:*
12:	gb_gss:*
13:	em_gss_hum:*
14:	em_gss_inv:*
15:	em_gss_pln:*
16:	em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	984.6	29.0	987	9	AL551235
2	969.8	28.6	1039	9	AL574873
3	795.2	23.4	1603	11	AK008530
4	596.2	17.6	763	10	BI913504
5	594.6	17.5	676	10	BI561069
6	581.8	17.1	635	10	BG722079
7	536.4	15.8	609	10	BI561359
8	502.8	14.8	766	10	BG962298
9	487.8	14.4	1154	10	BM460886
10	472	13.9	879	10	BG401683
11	469.4	13.8	965	10	BF789159
12	458.8	13.5	462	9	AW772472
13	451.2	13.3	686	9	BB652968
14	451	13.3	452	10	BG428060
15	440.6	13.0	447	9	AI628611
16	430.4	12.7	443	9	AA397955
17	423.6	12.5	437	9	AI831883

c 18	422.8	12.4	426	9	AW271467
c 19	421	12.4	434	9	AA420969
c 20	420.8	12.4	555	9	AA162058
c 21	413.6	12.2	741	10	BG772384
c 22	413.4	12.2	778	10	BI831019
c 23	399	11.7	556	10	BM030353
c 24	384.2	11.3	398	10	BF091535
c 25	384	11.3	386	9	AA416585
c 26	382.6	11.3	842	10	BF784147
c 27	356.6	10.5	492	9	AW260204
c 28	353	10.4	353	9	AI928360
c 29	349.4	10.3	351	9	AW590883
c 30	348	10.2	574	9	AI116021
c 31	344	10.1	347	9	AW848579
c 32	343.4	10.1	348	9	AL039738
c 33	342.4	10.1	471	9	AW358360
c 34	320.6	9.4	392	10	BF334479
c 35	318.8	9.4	459	9	BE138336
c 36	314.4	9.3	552	9	AI036772
c 37	312.6	9.2	432	9	AW258861
c 38	306	9.0	418	9	BB687602
c 39	301	8.9	301	9	AA421125
c 40	298	8.8	299	10	BE843196
c 41	284.2	8.4	501	10	BM445902
c 42	277.8	8.2	451	10	BE722366
c 43	257.2	7.6	312	9	AW001638
c 44	256.8	7.6	292	10	BE843206
c 45	247.8	7.3	426	9	AI035384

## ALIGNMENTS

RESULT 1  
AL551235  
LOCUS  
DEFINITION AL551235 LTI\_NFL006\_PL2 Homo sapiens cdna clone CS0DI041YM14 5 prime, mRNA sequence.  
ACCESSION AL551235  
VERSION AL551235.1 GI:12888986  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 987)  
AUTHORS Li, W.B., Gruber, C., Jesses, J. and Polayes, D.  
TITLE Full-length cdna libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES  
Location/Qualifiers  
1..987  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DI041YM14"  
/clone\_lib="LTI\_NFL006\_PL2"  
/tissue\_type="placenta"  
/note="Vector: pcwvSPORT 6; Site\_1: NotI; 1st strand cdna was primed with a NotI-oligo(dr) primer. Five prime end enriched, double-stranded cdna was digested with Not I and cloned into the Not I and Eco RV sites of the pcwvSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371  
Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"  
BASE COUNT 322 a 182 c 211 g 270 t 2 others  
ORIGIN

Query Match		29.0%; Score 984.6; DB 9; Length 987;
Best Local Similarity		99.7%; Pred. No. 1.3e-193;
Matches 984; Conservative		2; Mismatches 1; Indels 0; Gaps 0;
QY	1661	AAGCATTGTCAAGCAGCTAAACATGAAGGCCCTCTGCACAAATGTGACATCTCAAACT 1720
DB	1	AAGCATTGTCAAGCAGCTAAACATGAAGGCCCTCTGCACAAATGTGACATCTCAAACT 60
QY	1721	CTACAGAAGCTGGACAGAACTGTTCATATGCTGAGGCTTGAAATCAGAACCCCTGGA 1780
DB	61	CTACAGAAGCTGGACAGAACTGTTCATATGCTGAGGCTTGAAATCAGAACCCCTGGA 120
QY	1781	CCCTAGCATTGGAAATGTTGTAGGAGCAAGAACATGAATGTAAAGCCACTCTCAACT 1840
DB	121	CCCTAGCATTGGAAATGTTGTAGGAGCAAGAACATGAATGTAAAGCCACTCTCAACT 180
QY	1841	ACTTTAGCCCTTATTTACCTGGCTGAAAGACCAGCAAGAAATCTTTTGTGGGATGGA 1900
DB	181	ACTTTAGCCCTTATTTACCTGGCTGAAAGACCAGCAAGAAATCTTTTGTGGGATGGA 240
QY	1901	GTACCGACTGGAGTCCATATGACAGCAAGCATCAAGTGAGGATAAGCCCTAAATTCAG 1960
DB	241	GTACCGACTGGAGTCCATATGACAGCAAGCATCAAGTGAGGATAAGCCCTAAATTCAG 300
QY	1961	CTCTTGAGATAAAGCATATGAATGGAAACGACAAATGAATGTACCTGTCCGATCATCTG 2020
DB	301	CTCTTGAGATAAAGCATATGAATGGAAACGACAAATGAATGTACCTGTCCGATCATCTG 360
QY	2021	TTGCATATGCTATGAGCAGTACTTTTAAAGTAAAAATCAGATGATCTTTTGGGG 2080
DB	361	TTGCATATGCTATGAGCAGTACTTTTAAAGTAAAAATCAGATGATCTTTTGGGG 420
QY	2081	AGGAGATGCGAGTGGCTTAATTTGAACCAAGAAATCTCCTTTAATTTCTTTGTCACTG 2140
DB	421	AGGAGATGCGAGTGGCTTAATTTGAACCAAGAAATCTCCTTTAATTTCTTTGTCACTG 480
QY	2141	CACCTAAAAATGTCTGATATCATCTCTAGNACTGAAGTTCAGAAAGGCATCAGATGT 2200
DB	481	CACCTAAAAATGTCTGATATCATCTCTAGNACTGAAGTTCAGAAAGGCATCAGATGT 540
QY	2201	CCGGAGCCGTATCAATGATGCTTTCCGTCTGAATGACAAACAGCCTAGAGTTTCTGGGA 2260
DB	541	CCGGAGCCGTATCAATGATGCTTTCCGTCTGAATGACAAACAGCCTAGAGTTTCTGGGA 600
QY	2261	TACAGCCACACTTGGACCTCTCAACAGACCCCTGTTTCCATATGCTGATCTTCTGAGT 2320
DB	601	TACAGCCACACTTGGACCTCTCAACAGACCCCTGTTTCCATATGCTGATCTTCTGAGT 660
QY	2321	GAGTTGTATGGAGTATAGTGTGCTTGGCATCTGCTGATCTTCTGATCTTCTGAGT 2380
DB	661	GAGTTGTATGGAGTATAGTGTGCTTGGCATCTGCTGATCTTCTGATCTTCTGAGT 720
QY	2381	ATCGGAAGAGAAAAATAAGCAAGAGTGGAGAAAAATCCTTATGCCCTCCATCGATATTA 2440
DB	721	ATCGGAAGAGAAAAATAAGCAAGAGTGGAGAAAAATCCTTATGCCCTCCATCGATATTA 780
QY	2441	GCAAGGAGAAAAATAATCCAGGATTCCAAAACACTGATGATGTTCCAGACCTCTTTTGA 2500
DB	781	GCAAGGAGAAAAATAATCCAGGATTCCAAAACACTGATGATGTTCCAGACCTCTTTTGA 840
QY	2501	AAATCTATGTTTTTCTCTTGGAGTATTTTGTGTATGTAAATGTTAAATTTCTATGGA 2560
DB	841	AAATCTATGTTTTTCTCTTGGAGTATTTTGTGTATGTAAATGTTAAATTTCTATGGA 900
QY	2561	TAGAAAAATAAGATGATAAGATATCATTTAAATGTCAAAACTATGACTCTGTTTCAGAA 2620
DB	901	TAGAAAAATAAGATGATAAGATATCATTTAAATGTCAAAACTATGACTCTGTTTCAGAA 960
QY	2621	AAAAATTTGCCAAGACAAATGCCCCA 2647
DB	961	AAAAATTTGCCAAGACAAATGCCCCA 987
RESULT 2		
AL574873/c		
LOCUS		AL574873 LTI_NFL006_PL2 1039 bp mRNA linear EST 16-FEB-2001
DEFINITION		prime, mRNA sequence.
ACCESSION		AL574873
VERSION		AL574873.1 GI:12935492
KEYWORDS		EST.
SOURCE		human.
ORGANISM		Homo sapiens
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS		Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE		Full-length cDNA libraries and normalization
JOURNAL		Unpublished (2001)
COMMENT		Contact: Genoscope
Genoscope - Centre National de Sequencage		
BP 191 91006 EVRY cedex - France		
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr.		
FEATURES		
source		1..1039
Location/Qualifiers		
/organism="Homo sapiens"		
/db_xref="taxon:9606"		
/clone_lib="CS0DI041YM14"		
/tissue_type="placenta"		
/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA		
was primed with a NotI-oligo(dT) primer. Five prime end		
enriched, double-stranded cDNA was digested with Not I and		
cloned into the Not I and Eco RV sites of the pCMVSPORT 6		
vector. Library was normalized. Library was constructed by		
Life Technologies. Contact : Feng Liang Life Technologies,		
a division of Invitrogen 9800 Medical Center Drive		
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371		
Email : fliang@lifetech.com URL :		
http://fulllength.invitrogen.com		
BASE COUNT		328 a 224 c 177 g 303 t
ORIGIN		7 others
Query Match		28.6%; Score 969.8; DB 9; Length 1039;
Best Local Similarity		98.4%; Pred. No. 1.5e-190;
Matches 1026; Conservative		3; Mismatches 9; Indels 5; Gaps 5;
QY	2281	CCTAACACAGCCCTGTTTCCATATGCGCTGATTTGTTTGGAGTTGTGATGGAGTGATA 2340
DB	1039	CCTAACACAGCCCTG-TTCCMTATGCTGATG-TTTTGGAGTTGTGATGGAGTGATA 982
QY	2341	GTGGTTGGCATTTGTCATCTGATCTTCACTGGGATCAGAGATCGGAAGAAAAATAA 2400
DB	981	GTGGTTGGCATTTGTCATCTGATCTTAACTGGCATCAGACATCGGAAGAAAAATAA 922
QY	2401	GCAAGAACTGGAGAAAAATCCTTATGCTCCATCGATATTAGCAAGAGAAAAATAATCCA 2460
DB	921	GCAAGAACTGGAGAAAAATCCTTATGCTCCATCGATATTAGCAAGAGAAAAATAATCCA 862
QY	2461	GGATTCCAAAACACTGATGATGTTTCAGACCTCTTTTAGAAAAATCTATGTTTCCCTCT 2520
DB	861	GGATTCCAAAACACTGATGATGTTTCAGACCTCTTTTAGAAAAATCTATGTTTCCCTCT 802
QY	2521	TSAGTGATTTTGTGTTGTTAAATGTTAAATTTATGTTTATGAAAAATATAAGATGATA 2580
DB	801	TSAGTGATTTTGTGTTGTTAAATGTTAAATTTATGTTTATGAAAAATATAAGATGATA 742
QY	2581	AGATATCAATTAATGTCAAACTATGACTCTGTTCAG-AAAAAAATGTCCTCAAGACAA 2639
DB	741	AGATATCAATTAATGTCAAACTATGACTCTGTTCAG-AAAAAAATGTCCTCAAGACAA 682
QY	2640	CATGCCAAGGAGAGAGCATCTTCATTTGACATTTGCTTCAAGTATTTTCTCTCTCTGG 2699
DB	681	CATGCCAAGGAGAGAGCATCTTCATTTGACATTTGCTTCAAGTATTTTCTCTCTCTGG 622
QY	2700	ATTTGACTTCTGTTCTGTTTCTTAATAAGGATTTTGTATTAGAGTATATTAGGGAAGTG 2759

```
Db 621 ATTTTACTTCTGTTCTGTTCTTAATAAGGATTTTCTATTAGAGTATATTAGGAAAGTG 562
QY 2760 TGTATTTGGTCTCACAGGCTGTTCAAGGTAATCTAAATGTAATGTCGTTGAAATTTCT 2819
Db 561 TGTATTTGGTCTCACAGGCTGTTCAAGGTAATCTAAATGTAATGTCGTTGAAATTTCT 502
QY 2820 GAAGTTGAAACAGGATATATCATTTGGAGCAAGTGTGGATCTTGTATGGAATATGGAT 2879
Db 501 GAAGTTGAAACAGGATATATCATTTGGAGCAAGTGTGGATCTTGTATGGAATATGGAT 442
QY 2880 GGATTCACCTTGTAAAGACAGTGCCTGGGAACCTGGTGTAGCTGCAAGGATTGAGAAATGGCAT 2939
Db 441 GGATTCACCTTGTAAAGACAGTGCCTGGTACTGTTGTAGCTGCAAGGATTGAGAAATGGCAT 382
QY 2940 GCATTAGCTCACTTTCAATTTAATCCATTGTCAAGGATGACATGCTTTCTTCACAGTAAC 2999
Db 381 GCATTAGCTCACTTTCAATTTAATCCATTGTCAAGGATGACATGCTTTCTTCACAGTNACT 322
QY 3000 CAGTTCAAGTACTATGGTGTATTTGCTACAGTGTATTTGGAAATCGATCATGCTTTCTTC 3059
Db 321 CAGTTCAAGTACTATGGTGTATTTGCTACAGTGTATTTGGAAATCGATCATGCTTTCTTC 262
QY 3060 AAGGTGACAGCTTAAAGAGAGAAATCCAGGACAGGTAGAGGACATGCTTTTCA 3119
Db 261 AAGGTGACAGCTTAAAGAGAGAAATCCAGGACAGGTAGAGGACATGCTTTTCA 202
QY 3120 CTTCCAAGTCTGTGATCAACATCTCCCTGACAAACAAACCTAGAGCCAGGCGCCCTCCG 3179
Db 201 CTTCCAAGTCTGTGATCAACATCTCCCTGACAAACAAACCTAGAGCCAGGCGCCCTCCG 142
QY 3180 TGAATCCAGAGAGTGCCTGATGAAACTCATTTCTACTGTTCTCTAACTGTGGAGTGA 3239
Db 141 TGAATCCAGAGAGTGCCTGATGAAACTCATTTCTACTGTTCTCTAACTGTGGAGTGA 82
QY 3240 ATGGAATCCAACTGTATGTTCAACCTCTGAAGTGGGACCAGCTCTCTTAAATCTTTT 3299
Db 81 ATGGAATCCAACTGTATGTTCAACCTCTGAAGT-GGTACCAGCTCTCTTAAATCTTTT 23
QY 3300 GTATTTGCTCACAGTGTGAGC 3322
Db 22 GTA-TTGCTNACAGTGTGAGC 1

RESULT 3
AK008530
LOCUS
DEFINITION
Mus musculus adult male small intestine cDNA, RIKEN full-length
enriched library, clone:2010305L05:homolog to ANGIOTENSIN
CONVERTING ENZYME-LIKE PROTEIN (ACE-RELATED CARBOXYPEPTIDASE ACE2),
full insert sequence.
ACCESSION
AK008530
VERSION
AK008530.1 GI:12842766
KEYWORDS
Mus musculus (strain:C57BL/6J) adult male small intestine cDNA to
mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library
Clone:2010305L05.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (sites)
REFERENCE
Carninci,P., Shibata,Y., Hayatsu,N., Sugihara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
PUBLISHED
2 (sites)
REFERENCE
Carninci,P., Shibata,Y., Hayatsu,N., Sugihara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
MEDLINE
```

```
11042159
PUBMED
REFERENCE
AUTHORS
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Ikehata,T., Kashiwagi,K.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
PUBMED
MEDLINE
11076861
REFERENCE
AUTHORS
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5 (bases 1 to 1603)
ADACHI,J., AIZAWA,K., AKAHIRA,S., AKIMURA,T., AONO,H., ARAI,A.,
ARAKAWA,T., BALDARELLI,R., BONO,H., BROWNSTEIN,M., BULT,C.,
CARNINCI,P., FUKUDA,S., FUKUNISHI,Y., FURUNO,M., HANAGAKI,T.,
HARA,A., HAYATSU,N., HILL,D., HIRAMOTO,K., HIRAOKA,T., HORI,F.,
HUME,D., IMOTANI,K., ISHII,Y., ITOH,M., IZAWA,M., KASUKAWA,T.,
KATO,H., KAWAI,J., KOJIMA,Y., KONNO,H., KODA,M., KOYA,S.,
KURIHARA,C., MATSUYAMA,T., MIYAZAKI,A., NISHI,K., NOMURA,K.,
NUMAZAKI,R., OHNO,M., OKAZAKI,Y., OKIDO,T., OWA,C., QUACKENBUSH,J.,
SAITO,H., SAITO,R., SAKAI,K., SANO,H., SASAKI,D.,
SCHIRMI,L., SHIBATA,K., SHIBATA,Y., SHINAGAWA,A., SHIRAKI,T.,
SOGABE,Y., SUZUKI,H., TAGAMI,M., TAGAWA,A., TAKAHASHI,F.,
TANAKA,T., TEJIMA,Y., TOYA,T., YAMAMURA,T., YAMANAKA,I.,
YASUNISHI,A., YOSHIDA,K., YOSHINO,M., MURAMATSU,M. and
Hayashizaki,Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
FEATURES
Location/Qualifiers
1..1603
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="MGD:1902242"
/db_xref="taxon:10090"
/clone="2010305L05"
/sex="male"
/tissue_type="small intestine"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
49..798
/notes="data source:SFTR, source key:Q9NRA7, evidence:ISS
homolog to ANGIOTENSIN CONVERTING ENZYME-LIKE PROTEIN
(ACE-RELATED CARBOXYPEPTIDASE ACE2)
putative"
/codon_start=1
/protein_id="BAB25723.1"
/db_xref="GI:12842767"
/translaton="MLSLGNSEPTWKALENVVVGARNMDVKPLNLYFQPLFDWLKEQNR
NSFVGNTEWSPVADQSIKVRISLKSALGANAYETNNEMFLERSVAYAMRKYSII
KNQTPVPLEEDVRVSDIKPRVSFFVTSPQNSDVIPRSEVEDAIRMSRGRINDVFG
LDNLSLEFLGIHTPLEPPYQPTIWLIIIFGVNMAVYVVGIIITVTGIGKRRKNET
KRENSYDSMDTGKGSNAGFQNSDDAQTSF"
1584..1589
/notes="putative"
1603
/notes="putative"
polyA_signal
polyA_site
1603
BASE COUNT 465 a 321 c 366 g 450 t 1 others
```

ORIGIN

Query Match 23.4%; Score 795.2; DB 1.1; Length 1603;  
Best Local Similarity 72.6%; Pred. No. 2.1e-154;  
Matches 1203; Conservative 0; Mismatches 389; Indels 65; Gaps 11;

QY 1702 AAATGAGCATCTCAAACTCTACAGAGCTGGACAGAACTGTTCAATATGCTGAGGCTT 1761  
DB 1 AANTGTGACATCTCAATTCACATGAGCTGGGAGAGTTGCTCAAGATGCTGAGTCTT 60

QY 1762 GGAATATCAGACCCCTGGACCCCTAGCATTTGGAATATGTTAGGAGCAAGAACATGAAT 1821  
DB 61 GGAATATCAGACCCCTGGACCCCTGGACCCCTGGACCCCTGGACCCCTGGACCCCTGGAC 120

QY 1822 GTAAGCCACTGCTCAACTACTTGTGACCCCTTATTTACCTGGCTGGAAGACAGCAACAG 1881  
DB 121 GTAAGCCACTGCTCAACTACTTGTGACCCCTTATTTACCTGGCTGGAAGACAGCAACAG 180

QY 1882 AATCTCTTTTGGGGATGGAGTACCGACTGGAGTCCATATGCAGACCAAGACATCAAGTG 1941  
DB 181 AATCTCTTTTGGGGATGGAGTACCGACTGGAGTCCATATGCAGACCAAGACATCAAGTG 240

QY 1942 AGGATAAGCCTAAATCAGCTCTTGGAGATAAAGCATATGAATGGAAGCACAATGAAATG 2001  
DB 241 AGGATAAGCCTAAATCAGCTCTTGGAGCTAATGCATATGAATGGACCAACCAAGATG 300

QY 2002 TACCTGTCGGATCATCTGTTGGATATGCTATGAGGAGTACTTCTTTTAAAGTAAATAAT 2061  
DB 301 TACCTGTCGGATCATCTGTTGGATATGCTATGAGGAGTACTTCTTTTAAAGTAAATAAT 360

QY 2062 CAGATGATCTTTTGGGGAGGAGTGTGGAGTGGCTAATTTGAAACCAAGATCTCC 2121  
DB 361 CAGATGATCTTTTGGGGAGGAGTGTGGAGTGGCTAATTTGAAACCAAGATCTCC 420

QY 2122 TTAAATGTTCTTCTCAGCTGACCTTAAATATGCTGATATCATCTCCCTAGAACTGAAAT 2181  
DB 421 TTAAATGTTCTTCTCAGCTGACCTTAAATATGCTGATATCATCTCCCTAGAACTGAAAT 480

QY 2182 GAAAGGCCATCAGGATGTCGGGAGCCGTATCAATGATGCTTCCCTGATGACAC 2241  
DB 481 GAAAGGCCATCAGGATGTCGGGAGCCGTATCAATGATGCTTCCCTGATGACAC 540

QY 2242 AGCTAGAGTTCTGGGATACAGCCAACTGGACCTCTCAACAGCCCTGTTTCC 2301  
DB 541 AGCTAGAGTTCTGGGATACAGCCAACTGGACCTCTCAACAGCCCTGTTTCC 600

QY 2302 ATATGCTGATTTTGTGACCTGCTGATGGAGTGTAGTGGTGGCATTTGTCATCTG 2361  
DB 601 ATATGCTGATTTTGTGACCTGCTGATGGAGTGTAGTGGTGGCATTTGTCATCTG 660

QY 2362 ATCTTCACTGGGATCAGAGATCGGAGAGAAATTAAGCAAGAGTGGAGAAATCTCT 2421  
DB 661 ATCTTCACTGGGATCAGAGATCGGAGAGAAATTAAGCAAGAGTGGAGAAATCTCT 720

QY 2422 TATGCTCCATCATGATATTAGCAAGAGAGAAATTAATCAGGATTTCCAAACACTGATGAT 2481  
DB 721 TATGCTCCATCATGATATTAGCAAGAGAGAAATTAATCAGGATTTCCAAACACTGATGAT 780

QY 2482 GTTCAGACCTCTCTTTAGAAAAATCTATGTTTCTCTCTGAGTGTATTTCTGATGAT 2541  
DB 781 GTTCAGACCTCTCTTTAGAAAAATCTATGTTTCTCTCTGAGTGTATTTCTGATGAT 824

QY 2542 AATGTTAATTTTATGATATAGAAATATAGATGATTAAGATATCAATTAATGCAAAA 2601  
DB 825 AATGTTAATTTTATGATATAGAAATATAGATGATTAAGATATCAATTAATGCAAAA 884

QY 2602 CTATGACTCTGTTGAGAAAAATTTGTTCCAAAGACATCTGGCCAAAGAGAGATCT 2661  
DB 885 CTATGA---TCTGTTTCAAGTAACTGTTCCAAAG---AGCATCAGACTTGGAGGACATCT 939

QY 2662 TCATTGACATGCTTTTCAAGTATTTTCTCTCTCTGATTTGATTTCTGTTCTGTTCT 2721  
DB 940 TCATTGACATGCTTTTCAAGTATTTTCTCTCTCTGATTTGATTTCTGTTCTGTTCT 999

QY 2722 TAATAAGGATTTTGTATTAGATATATTAGGAAAGTGTGTTATTGTTCTCACAGGCTGT 2781  
DB 1000 TAATAAGGATTTTGTATTAGATATATTAGGAAAGTGTGTTATTGTTCTCACAGGCTGT 1059

QY 2782 TCAGGGATAATCTAAATTAATGTTGTAATTTCTGAAGTTGAAAACAAGGATATAT 2841  
DB 1060 CCAGGG-----TGATATGTTGGTAACTGGAGTTAGAAGATGAGATGATGCTCT 1108

QY 2842 CATTGGACCAAGTGTGGATCTTGTATGGAATATGATGGATCACTTGTAAAGGACAGTGC 2901  
DB 1109 CTTGGGGCAAGTGTGGCTTTCGGTGTGGCATCTG-----GG 1145

QY 2902 CTGGGAATCTGTAGCTGCAAGGATGGAATGGCATGATGATGCTTCTTCAATTAA 2961

DB 1146 CTGGAATCTGTGGGACTGTTGAGGTGGAATGGTGGCTGGTGCCTTCAATTCGAATCCAAG 1205

QY 2962 TCCATTGTCAAGG-ATGACATGCTTTCACAGTAACTCAAGTCAAGTCAAGTACTATGGTGTAT 3020

DB 1206 TGTACGTCATGCTCTGTGGCTTCTGCTTTCACACTTCTCACTTCAAGTACTGATGAAT 1265

QY 3021 TTGCTTACAGTATGTTTGGATCGATCATGCTTCTTCAAGTGCACAGTCTAAAGAGA 3080

DB 1266 TTGTTTCACAGTAATCTTGAATGGACTGCTCCCTTCTTGGAGTGCAGTTCACAGGAGA 1325

QY 3081 GAAGAATCCAGGGAACAGGTAGAGACATGCTTTTTCACACTTCTCAAGTGTCTGATCAAC 3140

DB 1326 -AAGAAGCCAGACATCAGGTAGAGACCATGACCTTTTCTCTTCTCCAAA---CTTGATCAAC 1381

QY 3141 ATCTCCCTGCAACACAAACATAGAGCCAGGCGCTCCGTTAACTCCGAGAGCATGCTG 3200

DB 1382 ATCTCTTCAACAGACACAGCTAG---CACAGGAATCCACGAACCCAGAGCATGCTG 1438

QY 3201 ATAGAACTCATTTTCTACTGTTTCTTACTGTTGAGTGTGAGTGAATGGAATTCGAATGTATGT 3260

DB 1439 TCAGAACT-ACTTCCATTATTCCTCCATTTGGAGTGAAGGAAATTCAGATGAATGC 1497

QY 3261 TCACCTCTGAGTGGGTACCCAGTCTCTTAATCTTTTGTATTTGCTCAGAGTGTGGA 3320

DB 1498 TCGATCTGTGAGTGGGTGCCAGTCTCTGAAATCTGTTGTTTCTCAGAGGCTGGA 1557

QY 3321 GCAGTGTGAGCACAAGCAGACACTCAATAATGCT 3357

DB 1558 GCAATGTTGAACACAAAGCCGAC-CTCAATAATAACT 1593

RESULT 4

BI913504

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

603179133F1 NIH\_MGC\_121 Homo sapiens cDNA clone IMAGE:5243048 5',  
mRNA sequence.  
BI913504.1 GI:16177911  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 763)  
NIH-MGC <http://mgi.nhl.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM11613 row: b column: 09  
High quality sequence stop: 675.

```
FEATURES
source
Location/Qualifiers
1..763
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:524048"
/clone_lib="NIH_MGC_121"
/lab_host="DH10B"
/Note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: EcoRV (destroyed); RNA source anonymous pool of 3
fetal brains, female age 20 weeks, female age 24 weeks,
and male age 26 weeks. Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range
0.7-3.5 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 017. Note:
this is a NIH_MGC Library."
BASE COUNT 237 a 167 c 170 g 189 t
ORIGIN
Query Match 17.6%; Score 596.2; DB 10; Length 763;
Best Local Similarity 94.2%; Pred. No. 3.5e-113;
Matches 719; Conservative 0; Mismatches 28; Indels 16; Gaps 9;
QY 64 CTTGGCTCAGAGGGGACGATGTCAGCTCTTCTGGCTCCTTCTCGCCTGTGTGCTGTA 123
Db 1 CTTGGCTCAGAGGGGACGATGTCAGCTCTTCTGGCTCCTTCTCGCCTGTGTGCTGTA 60
QY 124 ACTGCTGCTAGTCCACCAT-TGAGGACAGGCGACAGACATTTTGGACAGTTAACCA 182
Db 61 ACTGCTGCTAGTCCACCATGAGGACAGGCGACAGACATTTTGGACAGTTAACCA 120
QY 183 CGAAGCGGAGACCTGTTCTATCAAAAGTTCAGTCTTCTTGGAAATATTAACACCAATAT 242
Db 121 CGAAGCGGAGACCTGTTCTATCAAAAGTTCAGTCTTCTTGGAAATATTAACACCAATAT 180
QY 243 TACTGAAGAGAAATGCCAAACATGAATTAATGCTGGGACAAATGCTGCTCCCTTTTAAA 302
Db 181 TACTGAAGAGAAATGCCAAACATGAATTAATGCTGGGACAAATGCTGCTCCCTTTTAAA 240
QY 303 GGAACAGTCCACACTTGGCCAAATGTATCCACTACAGAAATTCAGAAATCTCAGAGTCAA 362
Db 241 GGAACAGTCCACACTTGGCCAAATGTATCCACTACAGAAATTCAGAAATCTCAGAGTCAA 300
QY 363 GCTTCAGCTCAGGCTCTTCAGCAAAATGGGTCTTCAGTGTCTCAGAGACAGAGCAA 422
Db 301 GCTTCAGCTCAGGCTCTTCAGCAAAATGGGTCTTCAGTGTCTCAGAGACAGAGCAA 360
QY 423 ACGTTGAACACAATTTCAAATACAAATGAGCACCACCTCTACAGTACTGGAAAGATTTGTAA 482
Db 361 ACGTTGAACACAATTTCAAATACAAATGAGCACCACCTCTACAGTACTGGAAAGATTTGTAA 420
QY 483 CCCAGATAATCCAGAAATGCTTATTACTTGAACAGGTTTGAATGAATAATATGCGAAA 542
Db 421 CCCAGATAATCCAGAAATGCTTATTACTTGAACAGGTTTGAATGAATAATATGCGAAA 480
QY 543 CAGTTTAGACTACAAATGAGAGCTCTGGCTTGGGAAAGC---TGGAGATCTGAGTCTGG 599
Db 481 CAGTTTAGACTACAAATGAGAGCTCTGGCTTGGGAAAGC---TGGAGATCTGAGTCTGG 540
QY 600 CAAGCAGCTGAGGCAATATATGAAGAGTATGTGTCTTGAANA-ATGAGATGCGAAGAG 658
Db 541 CAAGCAGCTGAGGCAATATATGAAGAGTATGTGTCTTGAANAATGAGATGCGAAGAG 600
QY 659 CAATCATATTAGGAGCTATGGGATTTATGGAG-AGGAGACTATGAAGT--AAATGGGG 715
Db 601 CAATCATATTAGGAGCTATGGGATTTATGGAGAGGAGAGACTATGAAGTACAAATGGGGT 660
QY 716 TAGATGCTATGACTACAGCCGCGG-CCAGTTTGATTGAA---CATGTGGAACATACCTTT 771
Db 661 AGCATGCTATGACTACCGCCGCGGAGTTGATTCGAACGACTCTGGACATACCTTT 720
QY 772 ---GAAGAGATTAAACCAATTATA-TGAACATCTTCATGCTAT 810
```

```
Db 721 TGACGAGACTAAACCCATTATATCTGAACCTCTTCTATGCTAT 763
|||||
RESULT 5
BI561069 603254275F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5296531 5',
LOCUS mRNA sequence.
DEFINITION BI561069 676 bp mRNA linear EST 05-SEP-2001
ACCESSION BI561069
VERSION BI561069.1 GI:15448383
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 676)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs.femail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11750 row: f column: 20
High quality sequence stop: 674.
FEATURES
Location/Qualifiers
1..676
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5296531"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/Note="Organ: testis; Vector: pBluescriptR (modified
pluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.2 kb and
normalized to 10^5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 215 a 134 c 161 g 166 t
ORIGIN
Query Match 17.5%; Score 594.6; DB 10; Length 676;
Best Local Similarity 96.5%; Pred. No. 7.5e-113;
Matches 640; Conservative 0; Mismatches 19; Indels 4; Gaps 3;
QY 41 GAGCGCGCGCGCGCGGAGGATATCTTGGCTCAGAGGGAGCATGTCAAGCTCTTCTGGC 100
Db 14 GAAAGTCATTCAGTGGATGTATCTTGGCTCAGAGGGAGCATGTCAAGCTCTTCTGGC 73
QY 101 TCCTTCTCAGCCTTGTGCTGTAACCTGCTCAGTCCACCATTCAGGACAGGCAAGA 160
Db 74 TCCTTCTCAGCCTTGTGCTGTAACCTGCTCAGTCCACCATTCAGGACAGGCAAGA 133
QY 161 CATTTTGGCAAGTAAACACGAGCGAGACCTGTTCTATCAAGTTTCACTTGCTT 220
Db 134 CATTTTGGCAAGTAAACACGAGCGAGACCTGTTCTATCAAGTTTCACTTGCTT 193
QY 221 CTTGGAATTTAACAACAATATTACTGAAGAAGATGTCCAAACATGAATATGCTGGG 280
Db 194 CTTGGAATTTAACAACAATATTACTGAAGAAGATGTCCAAACATGAATATGCTGGG 253
QY 281 ACAATGCTGCTGCTTTTAAAGGAA-CAGTCCACACTTGGCCAAATGTATCCATACAA 339
|||||
```

```
Db 254 ACAAAATGGTCTGCCTTTTAAAGAGACAGCTCCACACTTGGCCCAAAATGATCCACTACAA 313
Qy 340 GAAATTCAGAAATCTCACAGTCAAGCTTCAGCTGAGGCTCTTTCAGCAAAATGGGCTCTTCA 399
Db 314 GAAATTCAGAAATCTCACAGTCAAGCTTCAGCTGAGGCTCTTTCAGCAAAATGGGCTCTTCA 373
Qy 400 GTGCTGTCAGAGACAGAGCAAAACGGTTGAACACAAATCTTAAATACAAATGAGCACCATC 459
Db 374 GTGCTCTCAGAGACAGAGCAAAACGGTTGAACACAAATCTTAAATACAAATGAGCACCATC 433
Qy 460 TACAGTACTGGAAAAGTTTGAACCCAGATAATCCACAAAGATGCTTATTACTTGAACCA 519
Db 434 TACAGTACTGGAAAAGTTTGAACCCAGATAATCCACAAAGATGCTTATTACTTGAACCA 493
Qy 520 GCTTTCAATCAATATGGAACACAGTTTACACTACAAATGAGAGGCTCTGGGCTTGGGAA 579
Db 494 GCTTTGAATCAATATGGAACACAGTTTACACTACAAATGAGAGGCTCTGGGCTTGGGAA 553
Qy 580 AGCTGAGATCTCAGGTCGGCAAGCAGCTGAGGCAATATATGAAGAGTATGTGCTTGG 639
Db 554 AGCTGAGATCTCAGGTCGGCAAGCAGCTGAGGCAATATATGAAGAGTATGTGCTTGG 613
Qy 640 -AAAAATGAGATGGCAAGACCAATCTATTAGG--ACTATGGGGATATTATGGAGAGCA 696
Db 614 AAAAATGAGATGGCAAGACCAATCTATTAGGAGGACTATGGGGATATTATGGAGAGCA 673
Qy 697 GAC 699
Db 674 GAC 676

RESULT 6
Bg722079
LOCUS
DEFINITION
Bg722079
ACCESSION
Bg722079
VERSION
Bg722079
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 635)
NIH-MGC
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10752 row: c column: 21
High quality sequence stop: 633.
Location/Qualifiers
1..635
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4830668"
/clone_lib="NIH-MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescript (modified
pBluescript KS+); Site.1: BamHI; Site.2: SalI-XhoI (gtcag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.2 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
```

```
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH-MGC Library.*
BASE COUNT 202 a 131 c 144 g 158 t
ORIGIN
Query Match 17.1%; Score 581.8; DB 10; Length 635;
Best Local Similarity 98.4%; Pred. No. 3.4e-110; Indels 3; Gaps 3;
Matches 619; Conservative 0; Mismatches 7;
Qy 51 CCGGGGAGGATCTTGGCTCACA-GGGGACGATGTCAGCTCTTGGCTCTCTCTCA 109
Db 9 CGGTGGATGATCTTGGCTCACAAGGGGAGGATGTCAGCTCTTGGCTCTCTCA 68
Qy 110 GCCTTGTGCTGTAAGTCTGCTCAGTCCAGCATTTAGGAACAGGCAAGCAATTTTGG 169
Db 69 GCCTTG-TGCTGTAAGTCTGCTCAGTCCAGCATTTAGGAACAGGCAAGCA-TTTTGG 126
Qy 170 ACAAGTTTAAACACAGGAGGAGGCTCTTCTATCAAAAGTTCACCTTCTTGGAAAT 229
Db 127 ACAAGTTTAAACACAGGAGGAGGCTCTTCTATCAAAAGTTCACCTTCTTGGAAAT 186
Qy 230 ATACACCAATATTACTGAAGAGATGTCCAAACATGAATATGCTGGGACAAATGGT 289
Db 187 ATACACCAATATTACTGAAGAGATGTCCAAACATGAATATGCTGGGACAAATGGT 246
Qy 290 CTGCTTTTAAAGGAACAGTCCACACTTGGCCAAATGTATCCACTACAAAGAAATTCAGA 349
Db 247 CTGCTTTTAAAGGAACAGTCCACACTTGGCCAAATGTATCCACTACAAAGAAATTCAGA 306
Qy 350 ATCTCACAGTCAAGCTTCAGCTGAGGCTCTTTCAGCAAAATGGCTCTTCAAGTCTGAC 409
Db 307 ATCTCACAGTCAAGCTTCAGCTGAGGCTCTTTCAGCAAAATGGCTCTTCAAGTCTGAC 366
Qy 410 AAGCAAGAGCAACGGTTGAACACAAATTTAAATACAAATGAGCACCATCTACAGTACTG 469
Db 367 AAGCAAGAGCAACGGTTGAACACAAATTTAAATACAAATGAGCACCATCTACAGTACTG 426
Qy 470 GAAAAGTTTGAACCCAGATAATCCACAAAGATGCTTATTACTTGAACAGGTTTGAATG 529
Db 427 GAAAAGTTTGAACCCAGATAATCCACAAAGATGCTTATTACTTGAACAGGTTTGAATG 486
Qy 530 AAATAATGGCAACAGTTTAGACTACAAATGAGAGGCTCTGGGCTTGGGAAGCTGGAGAT 589
Db 487 AAATAATGGCAACAGTTTAGACTACAAATGAGAGGCTCTGGGCTTGGGAAGCTGGAGAT 546
Qy 590 CTGAGTGGCAAGAGCAGCTGAGGCCATTATATGAAGAGTATGTGGCTTGAAGAAATGAGA 649
Db 547 CTGAGTGGCAAGAGCAGCTGAGGCCATTATATGAAGAGTATGTGGCTTGAAGAAATGAGA 606
Qy 650 TGGCAAGAGCAATCATTTATGAGGACTAT 678
Db 607 TGGCAAGAGCAATCATTTATGAGGACTAT 635

RESULT 7
Bg722079
LOCUS
DEFINITION
Bg722079
ACCESSION
Bg722079
VERSION
Bg722079
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 609)
NIH-MGC
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
```

cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM11752 row: j column: 05  
High quality sequence stop: 606.

FEATURES  
source  
1. 609  
location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5297380"  
/clone\_lib="NIH\_MGC\_97"  
/lab\_host="DH10B"  
/note="Organ: testis; Vector: pBluescriptR (modified  
pBluescript KS+); Site.1: BamHI; Site.2: SalI-XhoI (gtcgag  
) ; Oligo-4T primed using primer 5'-TTTTTTTTTTTTTTTNN-3',  
size-selected for average insert size 2.2 kb and  
normalized to R0T 5. This is a primary library enriched  
for full-length clones and constructed using the  
Cap-trapper method (Carninci, in preparation). Library  
constructed by M. Brownstein (NHGRI, National  
Institutes of Health). Note: this is a NIH\_MGC Library."  
BASE COUNT 193 a 128 c 137 g 151 t

Query Match 15.8%; Score 536.4; DB 10; Length 609;  
Best Local Similarity 97.5%; Pred. No. 8.5e-101;  
Matches 587; Conservative 0; Mismatches 11; Indels 4; Gaps 4;

QY 51 CCGGGGCGAGTATCTTGGCTCACAGGGGACGATGTCAAGCTCTTCTGGCTCCTCTTCAG 110  
| | | | |  
Db 9 CAGTGGATGTGATCTTGGCTCACAGGGGACGATGTCAAGCTCTTCTGGCTCCTCTTCAG 68  
| | | | |  
QY 111 CCTTGTGCTGTAAGTCTGCTAGTCCACCATGAGGAAACAGGCAACATTTTGGGA 170  
| | | | |  
Db 69 CCTATGTGCTGTAAGTCTGCTAGTCCACCATGAGGAAACAGGCAACATTTTGGGA 128  
| | | | |  
QY 171 CAAGTTTAAACCAAGCCGAGAGCTTCTTATCAAGTTTCACTTCTTGGGAATTA 230  
| | | | |  
Db 129 CAAGTTTAAACCAAGCCGAGAGCTTCTTATCAAGTTTCACTTCTTGGGAATTA 188  
| | | | |  
QY 231 TAACACCAATATTACTGAAGAGATGTCCAAACATGAATATGCTGGGACAAATGGTC 290  
| | | | |  
Db 189 TAACACCAATATTACTGAAGAGATGTCCAAACATGAATATGCTGGGACAAATGGTC 248  
| | | | |  
QY 291 TGGCTTTTAA-AAGGAACAGTCCACATTTGCCAAATGTATCCACTTACAGAAATTCAGA 349  
| | | | |  
Db 249 TGGCTTTTAAAGGAACAGTCCACATTTGCCAAATGTATCCACTTACAGAAATTCAGA 308  
| | | | |  
QY 350 ATCTCAGCTCAAGCTTCAGCTGAGGCTCTTCAGAA-AATGGGTCTTCAGTGTCTCA 408  
| | | | |  
Db 309 ATCTCAGCTCAAGCTTCAGCTGAGGCTCTTCAGAA-AATGGGTCTTCAGTGTCTCA 368  
| | | | |  
QY 409 GAAGACAAGAGCAACGGTTCAACAC-AATCTTAATCAATGAGGACCATCTACAGTAC 467  
| | | | |  
Db 369 GAAGACAAGAGCAACGGTTCAACACGAATTTCAATCAATGAGGACCATCTACAGTAC 428  
| | | | |  
QY 468 TGGAAAGATTGTAAACCCAGATAATCCAAAGAATGCTTATTTGAACCCAGTTTGA 527  
| | | | |  
Db 429 TGGAAAGATTGTAAACCCAGATAATCCAAAGAATGCTTATTTGAACCCAGTTTGA 488  
| | | | |  
QY 528 TGAATAATGGCAACAGTTTATAGACTACAATGAGAGGCTCTGGGCTTGGGAAAGCTGGAG 587  
| | | | |  
Db 489 TGAATAATGGCAACAGTTTATAGACTACAATGAGAGGCTCTGGGCTTGGGAAAGCTGGAG 548  
| | | | |  
QY 588 ATCTGAGGTGGCAAGCAGCTTGGGCCATTATATGAAGAGTATGTTGGTCTTGAATAATGA 647  
| | | | |  
Db 549 ATCTGAGGTGGG-AAGCAGCTGAGGCCATTATATCAAGAGTATGTTGGTCTTGAATAATGA 607  
| | | | |  
QY 648 GA 649

Db 608 GA 609

RESULT 8  
BG962298  
LOCUS  
DEFINITION  
BG962298 766 bp mRNA linear EST 12-JUN-2001  
608227090Fl NCI\_CGAP\_Co24 Mus musculus cDNA clone IMAGE:4981606 5',  
mRNA sequence.  
ACCESSION  
BG962298  
VERSION  
BG962298.1 GI:14349935  
KEYWORDS  
EST  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus  
REFERENCE  
1 (bases 1 to 766)  
AUTHORS  
NIH-MGC http://mgc.nci.nih.gov/  
JOURNAL  
National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10983 row: d column: 23  
High quality sequence stop: 708.

FEATURES  
source  
1. 766  
location/Qualifiers  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4981606"  
/clone\_lib="NCI\_CGAP\_Co24"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: colon; Vector: pCMV-SPORT6; Site.1: NotI;  
Site.2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 1.6 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."  
BASE COUNT 233 a 156 c 179 g 198 t

ORIGIN  
Query Match 14.8%; Score 502.8; DB 10; Length 766;  
Best Local Similarity 85.3%; Pred. No. 7.7e-94;  
Matches 631; Conservative 0; Mismatches 102; Indels 7; Gaps 6;  
QY 1327 AAGCATTTAAATCCATTTGCTCTCTCACCCTGATTTTCAAGAGACAATGAACAGAA 1386  
| | | | |  
Db 2 AAGCATCTGAATCCATTTGCTCTCTCACCCTGATTTTCAAGAGAGATAGCGAAACAGAG 61  
| | | | |  
QY 1387 ATAACTCTCTGCTCAACAGCACTCAGCATTTCTGGACTCTGCCATTTTACTTACATG 1446  
| | | | |  
Db 62 ATAACTCTCTGCTCAACAGCACTCAGCATTTCTGGACTCTGCCATTTTACTTACATG 121  
| | | | |  
QY 1447 TTAGAAAGTGGAGTGGATGGTCTTTAAAGGGGAAATTTCCAAAGACAGTGGATGAAA 1506  
| | | | |  
Db 122 TTAGAAAGTGGAGTGGATGGTCTTTCCGGGTGAATTTCCAAAGACAGTGGATGAAA 181  
| | | | |  
QY 1507 AAGTGGTGGAGATGAAGCGAGAGATAGTTGGGTGGTGGAACTGTGCCCATGATGAA 1566  
| | | | |  
Db 182 AAGTGGTGGAGATGAAGCGGAGATGCTTTGGTGGTGGAGCCCTCTGCCCTCATGATGA 241  
| | | | |  
QY 1567 ACATCTGTGACCCCGCATCTCTGTTCCATGTTTCTAATGATTAATCATTCATTCGATAT 1626  
| | | | |  
Db 242 ACATCTGTGACCCCTGCATCTCTGTTCCATGTTTCTAATGATTAATCATTCATTCGATAT 301  
| | | | |  
QY 1627 TACACAGGACCTTTTACCAATTTCCAGTTTCAAGAACACCTTTGTCAAGCAGCTTAACAT 1686  
| | | | |  
Db 302 TACACAGGACCTTTTACCAATTTCCAGTTTCAAGAACCTCTTTGTCAAGCAGCTAAGTAT 361  
| | | | |







REFERENCE  
1 (bases 1 to 879)  
AUTHORS NTH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: CLONTECH Laboratories, Inc.  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L10CM1335 row: d column: 13  
 High quality sequence stop: 619.

```

FEATURES
source
align quality sequence stop: 619.
Location/Qualifiers
1. .879
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4594140"
/clone_lib="NIH_MGC_75"
/lab_host="DH10B (11 phage-resistant)"
/note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site:1:
sfll (ggcgctcgccg); Site:2: sfll (ggccattatgcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGGCCATTATGGCC-3 and 3' adaptor sequence:
5'-ATTCTAGAGCCGAGCGCCGACATG-dh(30)BN-3' (Where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.65
kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
275 a 210 g 176 t 1 others
BASE COUNT

```

Query Match	Score 472;	DB 10;	Length 879;
Best Local Similarity	81.1%	Pred. No. 1.8e-87;	
Matches 674;	Conservative	0;	Mismatches 36;
			Indels 121;
			Gaps 6;

Qy	1242	ACAACTTTTCTGCTAAGAAATGGAGCTTAATGAAGGATTCATGAAGCTGTGTGGGAAAT	1301
Db		1 ACAACTTTTCTGCTAAGAAATGGAGCTTAATGAAGGATTCATGAAGCTGTGTGGGAAAT	60
Qy	1302	CATGTCACCTTCTGCGAGCCACACCTAAGCATTTAAATCCATTCGTCTGTGCACCGGA	1361
Db	61	CATGTCACCTTCTGCGAGCCACACCTAAGCATTTAAATCCATTCGTCTGTGCACCGGA	120
Qy	1362	TTTTCAAGAAAGACAAATGAAACAGAAATAAATCTGCTGCTCAAACAAGCACATCAGATTGT	1421
Db	121	TTTTCAAGAAAGACAAATGAAACAGAAATAAATCTGCTGCTCAAACAAGCACATCAGATTG	179
Qy	1422	TGGGACTCTGCCATTTACTTTACATGTTAGAGAACTGGAGGTGGATGTCCTTTAAAGGGGA	1481
Db	180	TGGGACTCTGCCATTTACTTTACATGTTAGAGAACTGGAGGTGGATGTCCTTTAAAGGGGA	239
Qy	1482	AAATTTCCAAAGACACAGTGGATGAAAGAGTGGTGGGAGATGAAGCGAGAGATAGTTGGGGT	1541
Db	240	AAATTTCCAAAGACACAGTGGATGAAAGAGTGGTGGGAGATGAA-----	281
Qy	1542	GGTGGAACTGTGCCCATGATGAACAATACTGTGACCCGCATCTCTGTCTCCATGTTTC	1601
Db	282	-----	281
Qy	1602	TAATGATTACTCATTCATTCGATATTCACAAGGACGCTTTACCAATTCGAGTTTCAAGA	1661
Db	282	-----GACGCTTTTACCAATTCGAGTTTCAAGA	308
Qy	1662	AGCATTGTGTCAGCAGCTAAACATGAAGGCCCTCTGCACAAATGTCACTCTCAAACTC	1721
Db	309	AGCATTGTGTCAGCAGCTAAACATGAAGGCCCTCTGCACAAATGTGACATCTCAAACTC	368
Qy	1722	TACAGAAGCTGGACAGAAACTGTTTCAATATGCTGAGGCTTGGAAATACGAAACCTGGAC	1781

Db	369	TACGAAGCTGGACAGAAACTGTTCAATATATGCTGAGGCTTGGAATAATCAGAAACCTGGAC	428
Qy	1782	CCTAGCATTGGAAAATGTTGAGGACAAAGACATGAATTAAGGGCACATGCTCAACTA	1841
Db	429	CCTAGCATTGGAAAATGTTGAGGACAAAGACATGAATTAAGGGCACATGCTCAACTA	488
Qy	1842	CTTTGAGCCCTTATTACCTGGCTGAANGACCAAGAAATTCWTTTGTGGGATGGAG	1901
Db	489	CTTTGAGCCCTTATTACCTGGCTGAANGACCAAGAAATTCWTTTGTGGGATGGAG	548
Qy	1902	TACCGACTGGAGTCCATATGCAG-ACCAAAGCATCAAGTGAGGATAAGCCTAAATCAG	1960
Db	549	TACCGACTGGAGTCCATATGCAGAACCAAGCATCAAGTGAGGATAAGCCTAA-AT	604
Qy	1961	CTCTCTGGAGATPAGCATATCAATGGACGACATGAATGCTTACCTCTCCGATCATCTG	2020
Db	605	CAGTCTGGAGATPAGCATATGACTTGGGA--CGACATGAATGTGCTGTNCCGATCATCTG	661
Qy	2021	TTGCATA-TGCTATGAGGCAGTACTTTTTAAAGTAAAAAATCAGATGATT	2070
Db	662	TGGCATATTGCTATTACGGACTACTTTTAAAGTAAACATCCATGATCTTT	712

RESULT 11					
BF789159					
LOCUS	BF789159	965 bp	mRNA	linear	EST 12-JAN-2001
DEFINITION	602105037F1	NCI_CGAP_Kid14	Mus musculus	cDNA clone	IMAGE:4223312
5' mRNA sequence.					

ACCESSION	BF789159
VERSION	BF789159.1
KEYWORDS	GI:12094195
EST.	
SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 965)  
NH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs@email.nih.gov](mailto:cgabbs@email.nih.gov)  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E Consortium/LLNL at:

http://image.llnl.gov  
Plate: LLAM9811 row: 1 column: 09  
High quality sequence stop: 676.

FEATURES	Location/Qualifiers
1. .965	/organism="Mus musculus"
	/strain="FVB/N"
	/db_xref="taxon:10090"
	/clone="IMAGE:4223312"
	/clone_lib="NCI_CGAP Kid14"
	/lab_host="DH10B (T1 phage-resistant)"
	/note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
	Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
	Average insert size 1.75 kb. Constructed by Life
	Technologies. Note: This is a NCI_CGAP Library.  "
BASE COUNT	286 a 213 c 253 t 1 others

Query Match 13.8%; Score 459.4; DB 10; Length 965;  
Best Local Similarity 78.9%; Pred. No. 6.4e-87;

Qy	62	ATCTTGGCTCAGGGGACGATGTC	CAAGCTCTTCTCGGCTCCTTCTCAGCC	TGTTGCTG	121
Db	1	ATCTTTGGCGCAGGGGAAAGATGTC <td>CAGCTCTCTCGGCTCCTTCTCAGCC</td> <td>-TGTTCGCTG</td> <td>59</td>	CAGCTCTCTCGGCTCCTTCTCAGCC	-TGTTCGCTG	59

QY 122 TAACTGCTGCTCAGTCCACCAATGAGAACAGGCGCAAGACATTTTGGACAGTTTAAACC 181  
DB 60 TTACTACTGCTCAGTCCCTCACCAGAGAAATGCCAGACATTTTAAACAACCTTAATC 119  
QY 182 AGGAAGCGAAGACCTGTCTATCAAAAGTTCACCTTCTTGGAGTATATACACACAATA 241  
DB 120 AGGAAGCTGAAGACCTGCTTATCAAAAGTTCACCTTCTTGGAGTATATATACATAACA 179  
QY 242 TTACTGAGAGAAATCTCCAAACATGAATATGCTGGGACAAATGCTGCTCTTTAA 301  
DB 180 TTACTGAGAAATGCCAAAGATGAGTGGCTGGAGCCAAATGCTGCTCTTTATG 239  
QY 302 AGGAACAGTCCACACTGCCCAAAATGTATCCACTACAAGAAATTCAGATCTCACAGTCA 361  
DB 240 AAGAACAGCTTAAGACTGCCCAAGTTCCTACTACAAGAAATCCAGACTCCGATCATCA 299  
QY 362 AGCTTCAGCTCAGCTCTTCAGCAAAATGGCTCTTCAGTGTCTGTCAGNAGACAGAGCA 421  
DB 300 AGCGTCAACTACAGCCCTTCAGCAAAATGGCTCTTCAGCACTCTCAGCAGCAAGAACA 359  
QY 422 AACGGTTGAACACAATTTAAATACAATGAGCACCATCTACAGTACTGGAAGTGTGTA 481  
DB 360 AACAGTTGACACAAATCTGACACCAATGAGCACCATTTACAGTACTGGAAGTGTGCA 419  
QY 482 ACCCAGATAATCCCAAGAAATGCTTATTTAGTGAACAGGTTTGAATGAATATATGGCAA 541  
DB 420 ACCCAAGAAACCCACAAAGATGCTTATTTAGTGAACAGGTTTGAATGAATATATGGCAA 479  
QY 542 ACAGTTTACAGTACATGAGAGGCTGCGCTTGGGAAAGCTGGAGATCTGAGGTCGGCA 601  
DB 480 CAAGCAGACATACAACTCTAGGCTGGGCAATGGGCTGGAGGCTGAGGTTGGCA 539  
QY 602 AGCAGTGAAGCCATTATATGAAGATATGCTGCTTGAAGAAATGAGATGGC-AAGAGCA 660  
DB 540 AGCAGTGAAGCCATTATATGAAGATATGCTGCTTGAAGAAATGAGATGGC-AAGAGCA 599  
QY 661 ATCATATTAGGACTATGGGATTTATGGAGAGAGACTATGAAGTAAATGGGTTAGAT 720  
DB 600 AACAATTATAGCATTGGGATTTATGGAGAGAGGACTATGAAGCAGAGGAGGACAGAT 659  
QY 721 GCTATGACTACAGCGCGGCGGCTGATTTGAAGATGTTGGAACATACCTTTGAAGAGATT 780  
DB 660 GTACAACATATAACCGGTAAACAGTATGAGATGTAAGTACTTCGC-----AGAA 712  
QY 781 AAACCATTTATGAACATCTTCATGCTATGAGGGCA 819  
DB 713 TCAGCATGGGTGAGCCCTCTTCATGCTATGAGGGGA 751

RESULT 12  
AN772472/c 462 bp mRNA linear EST 04-MAY-2000  
LOCUS hn74h04.x1 NCI\_CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:3033655 3',  
DEFINITION mRNA sequence.

ACCESSION AN772472  
VERSION AN772472.1 GI:7704538  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 462)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL, send email to:  
[info@image.llnl.gov](mailto:info@image.llnl.gov)  
Seq primer: -400P from Gibco  
High quality sequence stop: 453.  
Location/Qualifiers

FEATURES  
Source

1. 462  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3033655"  
/clone\_lib="NCI\_CGAP\_Kid11"  
/lab\_host="DH10B"

/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not 1; Site\_2: Eco RI;  
plasmid DNA from the normalized library NCI\_CGAP\_Kid3 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(cloneIDs 1322376-1323911, 1456007-1456775, and  
1500552-1502855). Subtraction by Bento Soares and M.  
Fatima Bernaldo."

BASE COUNT 132 a 98 c 100 g 132 t  
ORIGIN

Query Match 13.5%; Score 458.8; DB 9; Length 462;  
Best Local Similarity 99.6%; Pred. No. 1e-84;  
Matches 460; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2908 ACTGCTGAGTGCAGGATTCAGATGCGATGCGATAGCTCAGCTTCACTTAAATCCATT 2967  
DB 462 ACTGCTGAGTGCAGGATTCAGATGCGATGCGATAGCTCAGCTTCACTTAAATCCATT 403  
QY 2968 GTCAAGGATGACATGCTTCTTCCACAGTAACCTCAGTCAAGTACTATGCTGATTTGCCCTA 3027  
DB 402 GTCAAGGATGACATGCTTCTTCCACAGTAACCTCAGTCAAGTACTATGCTGATTTGCCCTA 343  
QY 3028 CAGTATGTTGGATCGATCATGCTTCTTCAAGGTGACAGGTCTAAAGAGAGAGAAAT 3087  
DB 342 CAGTATGTTGGATCGATCATGCTTCTTCAAGGTGACAGGTCTAAAGAGAGAGAAAT 283  
QY 3088 CCAGGGAACAGGTAGAGGACATGCTTCTTCACTTCCAAAGTCTTGTATCAACATCTCCC 3147  
DB 282 CCAGGGAACAGGTAGAGGACATGCTTCTTCACTTCCAAAGTCTTGTATCAACATCTCCC 223  
QY 3148 TGACAACAAACCTAGAGCCAGGGCCCTCCGTGAACCTCCAGAGAGATGCCGTATAGAAA 3207  
DB 222 TGACAACAAACCTAGAGCCAGGGCCCTCCGTGAACCTCCAGAGATGCCGTATAGAAA 163  
QY 3208 CTCATTCTACTGTTCTCTTAAGTGGAGTGAATGGAATTCCTCAACTGTATGTCACCCCT 3267  
DB 162 CTCATTCTACTGTTCTCTTAAGTGGAGTGAATGGAATTCCTCAACTGTATGTCACCCCT 103  
QY 3268 CTGAAGTGGTACCCAGTCTCTTAAATCTTTTGTATTTGCTCAGAGTGTGAGCAGTGC 3327  
DB 102 CTGAAGTGGTACCCAGTCTCTTAAATCTTTTGTATTTGCTCAGAGTGTGAGCAGTGC 43  
QY 3328 TGAGCACAAGCAGACACTCAATAATGCTAGATTTACACAC 3369  
DB 42 TGAGCACAAGCAGACACTCAATAATGCTAGATTTAAACAC 1

RESULT 13  
BB652968

LOCUS BB652968  
DEFINITION BB652968 RIKEN full-length enriched, adult male hippocampus Mus  
musculus cDNA clone C630041D11 5', mRNA sequence.

ACCESSION BB652968  
VERSION BB652968.1  
KEYWORDS BB652968.1 GI:15402926  
SOURCE EST.  
ORGANISM house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 686)  
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,  
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,  
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,  
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,  
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,  
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyota,T.,  
Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,  
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10  
(11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,  
Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa,  
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T.,  
Ishii,Y. and Hayashizaki,Y.  
Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.  
Funct. Genomics 2 pre, 172-L86 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.  
e mouse tissues.

FEATURES  
source

1. 686  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="C63004ID11"  
/clone.lib="RIKEN full-length enriched, adult male  
hippocampus"  
/sex="male"  
/tissue.type="hippocampus"  
/dev\_stage="adult"  
/lab\_host="SOLR"  
/note="Site 1: XhoI; Site 2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN, Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAAGGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5'  
GAGAGAGATTCGAGTTAATTAATTAATCCGCCGCCGCC 3']"  
211 a 151 c 169 g 151 t 4 others

BASE COUNT  
ORIGIN

Query Match 13.3% Score 451.2 DB 9; Length 686;  
Best Local Similarity 82.0% Pred No. 3.8e-83;  
Matches 530; Conservative 0; Mismatches 115; Indels 1; Gaps 1;  
QY 53 GGGCAGGTATCTTGGCTCACAGGGAGCATGTCAGAGCTCTTCTCTGCTCTCTCAGCC 112  
DB 3 GTGATGGATCTTGGCGCAGGGGAAAGATGTCAGCTCTCTCTGCTCTCTCTCAGCC 62  
QY 113 TTGTTGCTGTACTCTCTAGTCCACCATTCAGAGCAAGGCAAGACATTTTGGACA 172  
DB 63 TTGTTGCTGTACTCTCTAGTCCACCATTCAGAGCAAGGCAAGACATTTTAAACA 122  
QY 173 AGTTTAAACACAGACCGAAGACCTGTTCTATCAAAAGTTCACCTTCTTGGAAATATA 232  
DB 123 ACTTAAATCAGGAAGCTGAAGACCTGTCTTATCAAAAGTTCACCTTCTTGGAAATATA 182  
QY 233 ACACCAATATTACTGAAGAGATGTCACCAAAAGATGAATGCTGGGAGCAAAATGGTCTG 292  
DB 183 ATACTAACATTACTGAAGAGAAATGCCAAAAGATGAGTGGGTCGACCAAAATGGTCTG 242  
QY 293 CCTTTTAAAGCAAGCTCCACACTTGGCCAAATGTATCCACTACAGAAATTCAGAAATC 352  
DB 243 CCTTTTAAAGCAAGCTCCACACTTGGCCAAATGTATCCACTACAGAAATTCAGAAATC 302  
QY 353 TCACAGTCAAGCTTCAGCTGCAGGCTCTTCAGCAAAATGGTCTTCAGTCTCTCTCAGAA 412  
DB 303 CGATCATCAAGGCTCAACTACAGGCTTCAGCAAAATGGTCTTCAGCAACTCTCAGCAG 362  
QY 413 ACAAGAGCAAGCGTTGAACACAATTTAAATACAAATGAGCACCCTACAGTACTCGAA 472  
DB 363 ACAAGAGCAAGCGTTGAACACAATTTAAATACAAATGAGCACCCTACAGTACTCGAA 422  
QY 473 AAGTTTGAACCCAGATATCCACAGAAATGCTTATTTACTTGAACAGGTTTGAATGAAA 532  
DB 423 AAGTTTGAACCCAGATATCCACAGAAATGCTTATTTACTTGAACAGGTTTGAATGAAA 482  
QY 533 TAATGGCAAACTTTAGACTACAAATGAGAGGCTCTGGGCTTGGGAAAGCTGGAGATCTG 592  
DB 483 TAATGGCAAACTTTAGACTACAAATGAGAGGCTCTGGGCTTGGGAGGCTGGAGGCTG 542  
QY 593 AGTGGCAAGAGCTGAGGCTATATATGAGAGATATCTGGTCTTGAAGAAATGAGATG 651  
DB 543 AGTGGCAAGAGCTGAGGCTATATATGAGAGATATCTGGTCTTGAAGAAATGAGATG 602  
QY 652 GCAAGAGCAAAATCATTATGAGGCTATGGGATTTATGGAGAGGAG 697  
DB 603 GCAAGAGCAAAATCATTATGAGGCTATGGGATTTATGGAGAGGAG 648

RESULT 14  
BG428060  
LOCUS  
DEFINITION 602501471f1 NIH\_MGC\_75 Homo sapiens cDNA clone IMAGE:4615121 5',  
mRNA sequence.  
ACCESSION BG428060  
VERSION BG428060.1 GI:13334566  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 452)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaabbs-r@mail.nih.gov  
Tissue Procurement: CLONTECH Laboratories, Inc.  
cDNA Library Preparation: CLONTECH Laboratories, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov  
Plate: LLCM1367 row: n column: 18  
High quality sequence stop: 450.

FEATURES  
source  
1. .452  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4615121"  
/clone\_lib="NIH\_MGC\_75"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site\_1: SfiI (ggcccctggcc); Site\_2: SfiI (ggccattggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGCC-3' and 3' adaptor sequence: 5'-ATTAGAGCGCGGCGGACATG-dt(30)EN-3' (where B = A, C, or G, or T). Average insert size 1.65 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: This is a NIH\_MGC Library."  
BASE COUNT 144 a 81 c 103 g 124 t  
ORIGIN

Query Match 13.3%; Score 451; DB 10; Length 452;  
Best Local Similarity 100.0%; Pred. No. 4.1e-83;  
Matches 451; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1788 ATTGGAATTTGTAGAGCAAGCAACATGAATGTAAAGCCACTGCTCAACTACTTTGA 1847  
DB 1 ATTGGAATTTGTAGAGCAAGCAACATGAATGTAAAGCCACTGCTCAACTACTTTGA 60  
QY 1849 GCCCTATTACTGGCTGAAAGACAGAACAAAGAAATCTTTTGTGGGATGAGTACCGA 1907  
DB 61 GCCCTATTACTGGCTGAAAGACAGAACAAAGAAATCTTTTGTGGGATGAGTACCGA 120  
QY 1908 CTGGAGTCATATGAGACCAAGCATCAAGTGAAGCAAGCAAGCAAGCAAGCAAGCAAG 1967  
DB 121 CTGGAGTCATATGAGACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 180  
QY 1968 AGATAAGCATATGAATGGAACGACAAATGAAATGTACCTGTCCGATCATCTGTTCGATA 2027  
DB 181 AGATAAGCATATGAATGGAACGACAAATGAAATGTACCTGTCCGATCATCTGTTCGATA 240  
QY 2028 TGCTATGAGCGAGTACTTTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 2087  
DB 241 TGCTATGAGCGAGTACTTTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 300  
QY 2088 TGTGCGAGTGGCTAATTTGAAACCAAGATCTCCTTTAATTTTGTTCACATGCACCTAA 2147  
DB 301 TGTGCGAGTGGCTAATTTGAAACCAAGATCTCCTTTAATTTTGTTCACATGCACCTAA 360  
QY 2148 AAATGTGTGTATATCATCTCCAGACTGAAGTTGAAAGGCGCATCAGGATGTCGCGGAG 2207  
DB 361 AAATGTGTGTATATCATCTCCAGACTGAAGTTGAAAGGCGCATCAGGATGTCGCGGAG 420  
QY 2208 CCGTATCATATGATGCTTCCGCTGTAATGAC 2238  
DB 421 CCGTATCATATGATGCTTCCGCTGTAATGAC 451

RESULT 15  
LOCUS A1628611/c  
DEFINITION ty77b03.x1 NCI\_CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:2285069 3', mRNA linear EST 07-MAR-2000  
ACCESSION A1628611  
VERSION A1628611.1 GI:4665411  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE .1 (bases 1 to 447)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cyapbs-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/dbp/image/image.html  
Insert Length: 537 Std Error: 0.00  
Seq primer: -40UP from Gibco  
POLYA-No. Location/Qualifiers  
1. .447  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2285069"  
/clone\_lib="NCI\_CGAP\_Kid11"  
/lab\_host="DH10B"  
/note="Organ: kidney; Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; Plasmid DNA from the normalized library NCI\_CGAP\_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneids 132376-132391, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 127 a 89 c 97 g 134 t  
ORIGIN

Query Match 13.0%; Score 440.6; DB 9; Length 447;  
Best Local Similarity 99.1%; Pred. No. 5.9e-81;  
Matches 443; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2934 TGGCATGCAATAGCTCAGCTTCAATTTAATCCATTGTCAAGGATGACATGCTTTCTTCACA 2993  
DB 447 TGGCATGCAATAGCTCAGCTTCAATTTAATCCATTGTCAAGGATGACATGCTTTCTTCACA 388  
QY 2994 GTAACCTCAGTTCAGTACTATGGTGATTTGCCCTACAGTATGTTGGAATCGATCATGCT 3053  
DB 387 GTAACCTCAGTTCAGTACTATGGTGATTTGCCCTACAGTATGTTGGAATCGATCATGCT 328  
QY 3054 TTCTTCAAGGTGACAGGTCTAAAGAGAGAGAAATCCAGGAGACAGGTAGAGACATGCT 3113  
DB 327 TTCTTCAAGGTGACAGGTCTAAAGAGAGAGAAATCCAGGAGACAGGTAGAGACATGCT 268  
QY 3114 TTCTTCACTTCCAGGTGCTTGATCAACATCTCCCTCACAACACAAACATAGAGCCAGGGG 3173  
DB 267 TTCTTCACTTCCAGGTGCTTGATCAACATCTCCCTCACAACACAAACATAGAGCCAGGGG 208  
QY 3174 CCTCCGTGAACCTCCAGAGCATGCCGTGATAGAAACTCATTTCTACTGTTCTCTAATCTG 3233  
DB 207 CCTCCGTGAACCTCCAGAGCATGCCGTGATAGAAACTCATTTCTACTGTTCTCTAATCTG 148  
QY 3234 GAGTGAATGGAATTCACACTGTATGTTTCCACCTCTGAAGTGGTACCCAGTCTCTTAA 3293  
DB 147 GAGTGAATGGAATTCACACTGTATGTTTCCACCTCTGAAGTGGTACCCAGTCTCTTAA 88  
QY 3294 TCTTTTGTATTTGCTCACAGTGTGTTGAGCAGTGTGCTGAGCACAAGCAGACACTCAATAA 3353  
DB 87 TCTTTTGTATTTGCTCACAGTGTGTTGAGCAGTGTGCTGAGCACAAGCAGACACTCAATAA 28  
QY 3354 TGCTAGATTTACACACTCAAAAAA 3380  
DB 27 TGCTAGATTTACCCAAAAA 1

Thu Oct 10 09:35:25 2002  
.

Search completed: October 9, 2002, 15:30:24  
Job time : 2573 secs

us-09-635-501-1.rst

Page 13

